

AD-A062 151

DAVID W TAYLOR NAVAL SHIP RESEARCH AND DEVELOPMENT CE--ETC F/G 9/2
CHECK: A COMPUTER MODEL FOR ESTABLISHING DATA BOUNDS.(U)

DEC 78 G R HUMFELD
DTNSRDC-78/112

UNCLASSIFIED

NL

/ OF /
AD
A062 151



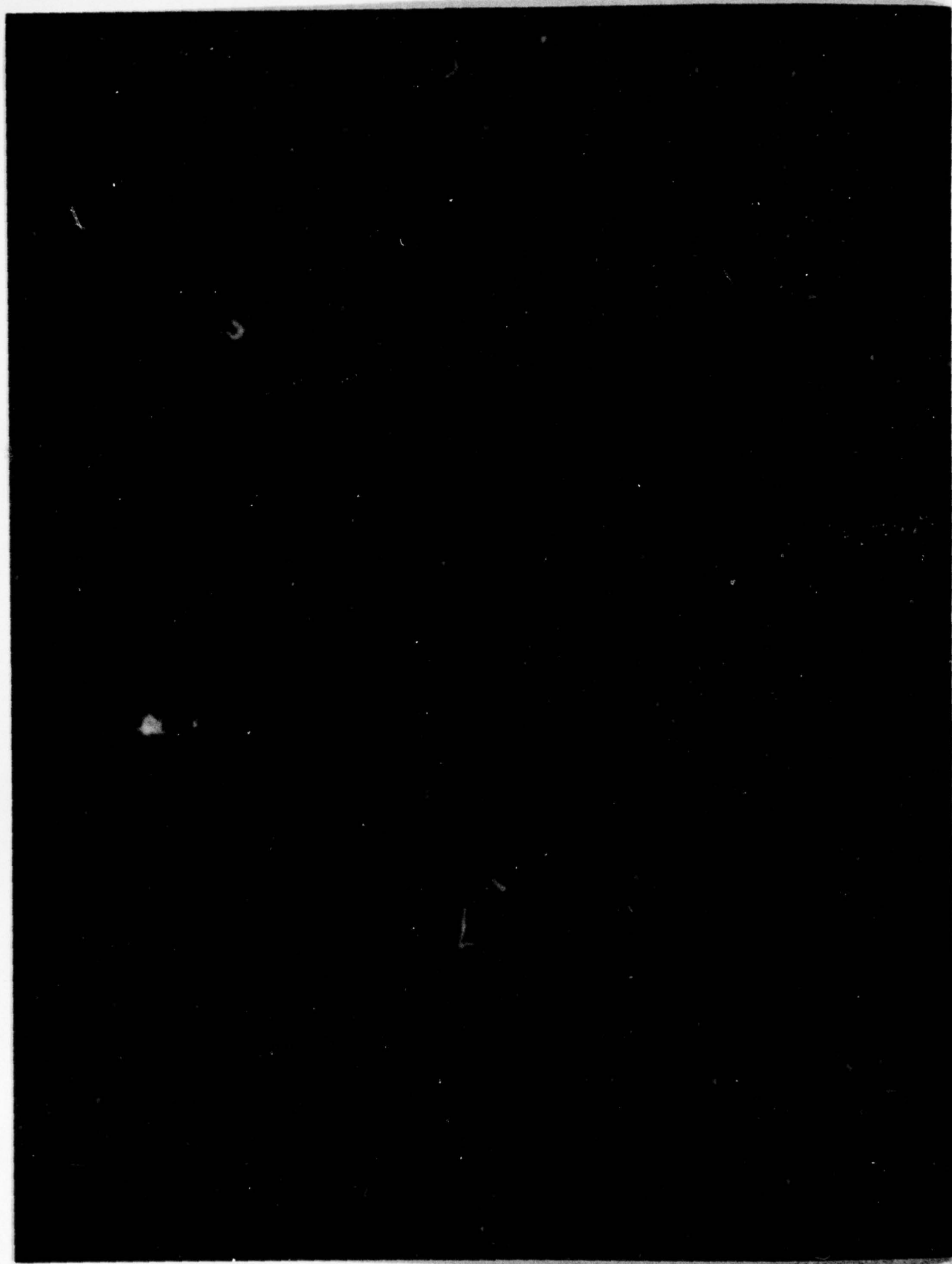
END
DATE
FILMED
3-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

DDC FILE COPY

ADA062151



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER DTNSRDC-78/112	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) <u>CHECK: A COMPUTER MODEL FOR ESTABLISHING DATA BOUNDS</u>		5. TYPE OF REPORT & PERIOD COVERED Final rept.
7. AUTHOR(s) George R. Humfeld		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS David W. Taylor Naval Ship Research and Development Center Bethesda, Maryland 20084		8. CONTRACT OR GRANT NUMBER(s) SEC PO 70298 SEC PO 80420
11. CONTROLLING OFFICE NAME AND ADDRESS <u>12 55 p.</u>		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (See reverse side)
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE Dec 1978
		13. NUMBER OF PAGES
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Data Analysis Statistical Distribution Data Bounds Probability Tolerance Limits Statistics Range Bounds Statistical Routines		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The Shipysest Maintenance Monitoring and Support Office (SMMSO) is establishing a new procedure whereby data collected by a SMMS Site Team will be typed into a DEC PDP-11/03 minicomputer at or near the data collection site. A route in the PDP-11 will check to see that each number entered lies within the bounds established for its data field. This report documents a FORTRAN computer model, called CHECK, which has been used to (Continued on reverse side)		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

387 682

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(Block 10)

Program Element 60000N

Task Area OMN

Work Unit 1870-003

Work Unit 1870-018

(Block 20 continued)

analyze existing data for each numerical field and to suggest sets of bounds for use in the PDP-11. Each data set examined by CHECK is tested for possible adherence to assumptions of exponentiality or normality. Bounds are suggested based on the results of these tests. In addition, CHECK indicates the expected probability that another data value from the same distribution will lie within the suggested bounds.

ACCESSION for	
NTIS	White Section <input checked="" type="checkbox"/>
DDC	Buff Section <input type="checkbox"/>
UNANNOUNCED	
JUSTIFICATION	
BY	DISPOSITION (AVAILABILITY)
A	

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iv
TABLE	iv
ABSTRACT.	1
ADMINISTRATIVE INFORMATION	1
1. INTRODUCTION	2
1.1 BACKGROUND	2
1.2 PURPOSE.	2
2. DESCRIPTION OF CHECK.	3
2.1 THE DATA READER.	3
2.2 THE EXPONENTIALITY CHECKER	3
2.2.1 A Large Sample Test	4
2.2.2 A Uniformity Test	4
2.2.3 A Two-Moment Test	5
2.3 THE EXPONENTIAL BOUNDS CALCULATOR.	5
2.3.1 Lower Limit Known	6
2.3.2 Lower Limit Unknown	7
2.3.3 Use Decision.	7
2.4 THE NORMALITY CHECKER	8
2.4.1 A Skewness Test	11
2.4.2 The Shapiro-Wilk Test	12
2.4.3 The Transformed D Statistic Test.	13
2.5 THE NORMAL BOUNDS CALCULATOR	14
2.5.1 Calculation of Bounds	14
2.5.2 Use Decision.	14
2.6 THE RANGE BOUNDS CALCULATOR	15
3. USE OF CHECK	19
3.1 INPUT	19
3.2 MECHANICS OF USE	19
3.2 OUTPUT	21
APPENDIX A: GLOSSARY	23
APPENDIX B: PROGRAM LISTING	25

	Page
APPENDIX C: SAMPLE RUN.	37
APPENDIX D: CROSS-REFERENCE TABLE OF VARIABLES AND PROGRAM USAGE	43
REFERENCES.	49

LIST OF FIGURES

1 - Graph of Equation (9)	9
2 - Graph $[1 - \exp([0.04 (75-n)]^{1/3} - 1.44224957)]R2/2$	10
3 - Properties of $P = \text{Prob}[X_1 \leq X \leq X_n]$	17

TABLE

1 - Detailed Summary of Figure 3.	18
---	----

ABSTRACT

✓ The Shipyards Maintenance Monitoring and Support Office (SMMSO) is establishing a new procedure whereby data collected by a SMMS Site Team will be typed into a DEC PDP-11/03 minicomputer at or near the data collection site. A routine in the PDP-11 will check to see that each number entered lies within the bounds established for its data field. This report documents a FORTRAN computer model, called CHECK, which has been used to analyze existing data for each numerical field and to suggest sets of bounds for use in the PDP-11. Each data set examined by CHECK is tested for possible adherence to assumptions of exponentiality or normality. Bounds are suggested based on the results of these tests. In addition, CHECK indicates the expected probability that another data value from the same distribution will lie within the suggested bounds.

ADMINISTRATIVE INFORMATION

This report is a result of work performed under Program Element
PE 60000N, Task Area OMN, Work Unit 1870-003, SEC PO 70298 and Work Unit
1870-018, SEC PO 80420 for NAVSEC 6107E.

1. INTRODUCTION

1.1 BACKGROUND

The Shipyards Maintenance Monitoring and Support Office (SMMSO) is establishing a new procedure whereby data collected by a SMMS Site Team (SST) will be typed into a DEC PDP-11/03 minicomputer (hereafter referred to as PDP-11) at or near the data collection site. A routine in the PDP-11 will check to see that each number typed in lies within bounds established for its data field. If any number does not lie within the bounds established, the PDP-11 will alert the SST member entering the data and allow him to reenter the number, leave the field blank until further testing can be conducted, or force the PDP-11 to accept the number as typed. Since the PDP-11 is on-site and the typing is to be done in the same time frame as the testing, many errors due to misreading measurement equipment, writing incorrect numbers on the Maintenance Requirement Cards (MRCs), and forgetting to perform tests or write down results can be corrected. The SST will often be able to return to the ship and conduct further testing if required. In addition, many keypunching errors will be avoided.

1.2 PURPOSE

A FORTRAN computer model called CHECK has been used on the CDC 6000 computers at the David W. Taylor Naval Ship Research and Development Center (DTNSRDC) to analyze existing data for each numeric field on the MRC's. Output of this model includes one or more suggested sets of bounds for use in the PDP-11. The results of using the model on the MRC numeric fields are available from SMMSO (NAVSEC 6107E1). This report provides documentation of the model.

2. DESCRIPTION OF CHECK

CHECK has six major parts: the data reader, the exponentiality checker, the exponential bounds calculator, the normality checker, the normal bounds calculator, and the range bounds calculator. Each part involving calculation of bounds also includes calculation of the estimated probability that some future measurement will also be found within the calculated bounds. The subroutine WSTEST is used by the exponentiality checker and the normality checker.

2.1 THE DATA READER

The data reader is designed to read data in a free format. The first thirty characters of each data record are read. Each character is then compared in its turn to fourteen "admissible" characters consisting of the ten digits, plus, minus, decimal point, and blank. The number is constructed as the identity of the characters read is determined by the comparisons. The comparisons end when the first of the following occurs:

- o All thirty characters have been determined
- o A blank is encountered after at least one digit has been found (this indicates that the complete number has been read)
- o An inadmissible character is encountered
- o Ten consecutive minus signs and decimal points have been encountered (this indicates the end of the data set).

Note that two or more numbers cannot be read from the same data record.

2.2 THE EXPONENTIALITY CHECKER

Epstein^{1*} described twelve tests for the validity of the exponential life assumption sometimes used in reliability tests. Two of these, a large sample test and a uniformity test, have been adapted for use in CHECK. A third test is used as a final selection criterion for inclusion of exponential bounds on a summary file, TAPE2 (see subsection 3.3).

*A complete listing of references is given on page 49.

78 12³ 11 003

2.2.1 A Large Sample Test

If X_1, X_2, \dots, X_n is a random sample from an exponential distribution arranged in ascending order (i.e., $X_j \leq X_{j+1}$), then a plot of X_j vs $\ln(n+1-j)$ can be fitted well by a straight line. As an indication of how well these points can be fitted with a straight line, CHECK calculates (in WSTEST) and prints the coefficient of determination of the least squares best linear fit to the data² as given by

$$R^2 = \frac{\left[\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \right]^2}{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (1)$$

where, for $i=1, 2, \dots, n$, $Y_i = \ln(n+1-i)$, and where \bar{X} is the average of the X_i 's and \bar{Y} is the average of Y_i 's. If R^2 is close to one (say, above 0.975), the fit is considered to be good. As Epstein¹ pointed out, at least fifty data values are required for this test to be meaningful.

2.2.2 A Uniformity Test

If X_1, X_2, \dots, X_n is a random sample (unordered) from an exponential distribution with probability density function (pdf)

$$f_X(x) = \begin{cases} (1/B)\exp[-(x-C)/B] & \text{for } x \geq C \\ 0 & \text{for } x < C \end{cases} \quad (2)$$

and if m is the largest integer no larger than $n/2$, then

$$f = \frac{(n-m) \sum_{i=1}^m (X_i - C)}{m \sum_{i=m+1}^n (X_i - C)} \quad (3)$$

has an F distribution with $2m$ and $2(n-m)$ degrees of freedom. CHECK uses the smallest of the X_i 's as an approximation to C , calculates f , and uses the International Mathematical and Statistical Libraries (IMSL)³ routine MDFD to find the significance level associated with f .

2.2.3 A Two-Moment Test

The standard deviation of a random variable having an exponential distribution with pdf given in Equation (2) is equal to the expected value of that random variable less C (see Johnson and Kotz,⁴ Chapter 18). The sample standard deviation of a random sample X_1, X_2, \dots, X_n is given by

$$s = \left[\sum_{i=1}^n (X_i - \bar{X})^2 / (n-1) \right]^{1/2} \quad (4)$$

If the sample mean minus the smallest sample value is less than half, or more than twice, the sample standard deviation, then exponentiality of the data must be doubted.

2.3 THE EXPONENTIAL BOUNDS CALCULATOR

If X is a random variable having an exponential distribution with pdf as given in Equation (2), then X can never take on a value smaller than C . Thus, C forms a natural lower bound for X . As will be seen in the next paragraph, $C+t(X-C)$ is a natural upper bound.

2.3.1 Lower Limit Known

If X_1, X_2, \dots, X_n is a random sample from an exponential distribution with pdf given in Equation (2), then each $(X_i - C)/B$ is known to have a gamma distribution with parameters 1, 1, and 0 (see Johnson and Kotz,⁴ Chapter 17, sections 1 and 5). Thus,

$$Y = n(\bar{X} - C)/B = \sum_{i=1}^n (X_i - C)/B \quad (5)$$

has a gamma distribution with parameters n , 1, and 0 (see Johnson and Kotz,⁴ Chapter 17, Section 2). That is, the pdf of Y is

$$f_Y(y) = \begin{cases} y^{n-1} e^{-y} / (n-1)! & \text{if } y \geq 0 \\ 0 & \text{if } y < 0 \end{cases} \quad (6)$$

If X is a random variable with pdf given by Equation (2), then

$$\begin{aligned} \text{Prob}[X > C + t(\bar{X} - C)] &= \text{Prob}[X > C + (tBY/n)] \\ &= \int_0^\infty \text{Prob}[X > C + (tBy/n) | Y=y] dF_Y(y) \\ &= \int_0^\infty \exp(-[C + (tBy/n) - C]/B) y^{n-1} e^{-y} dy / (n-1)! \\ &= \int_0^\infty \exp(-y[1 + (t/n)]) y^{n-1} dy / (n-1)! \\ &= \int_0^\infty z^{n-1} e^{-z} dz / [(n-1)! [1 + (t/n)]^n] \\ &= [1 + (t/n)]^{-n} \end{aligned} \quad (7)$$

since Equation (6) is a pdf (and, therefore, integrates to one). From Equation (7) it is seen that

$$\text{Prob}[X \leq C + t(\bar{X} - C)] = 1 - [1 + (t/n)]^{-n} \quad (8)$$

Note that once \bar{X} has been calculated, the probability that another independent draw from the exponential distribution is no larger than $C+t(\bar{X}-C)$ depends only on the value chosen for t . In particular, it is independent of the parameter B .

2.3.2 Lower Limit Unknown

Unfortunately, the value of the parameter C is also unknown in most cases. Johnson and Kotz⁴ (Chapter 18, section 5) suggest the smallest of X_1, X_2, \dots, X_n as an estimate for C . If this value is denoted by $L = C+a$ (a must be positive by the nature of the exponential distribution), then

$$\begin{aligned} \text{Prob}[X \leq L+t(\bar{X}-L)] &= \text{Prob}[X \leq C+t(\bar{X}-C)+a(1-t)] \\ &\leq \text{Prob}[X \leq C+t(\bar{X}-C)] = 1 - [1+(t/n)]^{-n} \end{aligned} \quad (9)$$

so long as $t > 1$. In order for Equation (9) to provide a "conservative" estimate, the inequality would have to be reversed. However, since the reason for the inequality is the difference between L and C , the right hand side of Equation (9) provides a good estimate so long as the physical situation prevents the data values from dropping below L .

2.3.3 Use Decision

This paragraph describes the criteria used in CHECK to decide whether or not a set of bounds based on an exponential assumption is to be recommended, and if so, which set. Any recommended bounds would be included on TAPE2 (see subsection 3.3). The criteria presented here are based on the professional judgment of the author and the requirements of the application for which CHECK was first written.

CHECK contains a variable, PCIND, whose value determines whether exponential bounds are to be recommended. The value of PCIND is initially set according to the result of the large sample test (see paragraph 2.2.1). If R_2 of Equation (1) is less than

$$FN = \begin{cases} 0 & \text{for } n \leq 5 \\ 0.30255 \ln n - 1.8237537 n^{0.09} + 2.2946 & \text{for } 5 < n < 397 \\ 0.98 & \text{for } n \geq 397 \end{cases} \quad (10)$$

PCIND is set equal to minus one, ultimately rejecting a recommendation of exponential bounds. If R_2 is larger than or equal to FN , the initial value of PCIND is initially set according to

$$PCIND = (1 - e^{[0.04(75-n)]^{1/3} - 1.44224952}) R_2 / 2 + 0.05(R_2 - FN) / (1 - FN) \quad (11)$$

Figure 1 is a plot of FN (Equation (10)). Figure 2 is a plot of the first term of Equation (11). This figure indicates that, for R_2 close to one, PCIND will start out close to $1/2$ for large sample sizes.

Next, PCIND is increased by the significance level of the uniformity test (see paragraph 2.2.2). Finally, PCIND is set equal to zero if the sample standard deviation (s of Equation (4)) is more than 1.8 times $X-L$ (see Equation (9)), or if $X-L$ is more than 1.8 times s (see paragraph 2.2.3). Exponential bounds are recommended if PCIND is at least $1/2$. The recommended bounds will be such that the expected probability is at least 0.9775 that another independent draw from the same presumed exponential distribution will lie within the bounds.

2.4 THE NORMALITY CHECKER

The normality checker uses a skewness test, the Shapiro-Wilk test (for samples of size fifty or less), and a transformed D statistic test (for samples of size larger than fifty) to test for normality. Shapiro et al.,⁵ in a comparative study of nine different tests, found that the Shapiro-Wilk test was the best overall test for normality, particularly for small sample sizes, but that the skewness test was much better against skewed alternatives. The transformed D statistic test used here was introduced by D'Agostino⁶ several years later and, therefore, was not included among the nine tests compared by Shapiro et al.

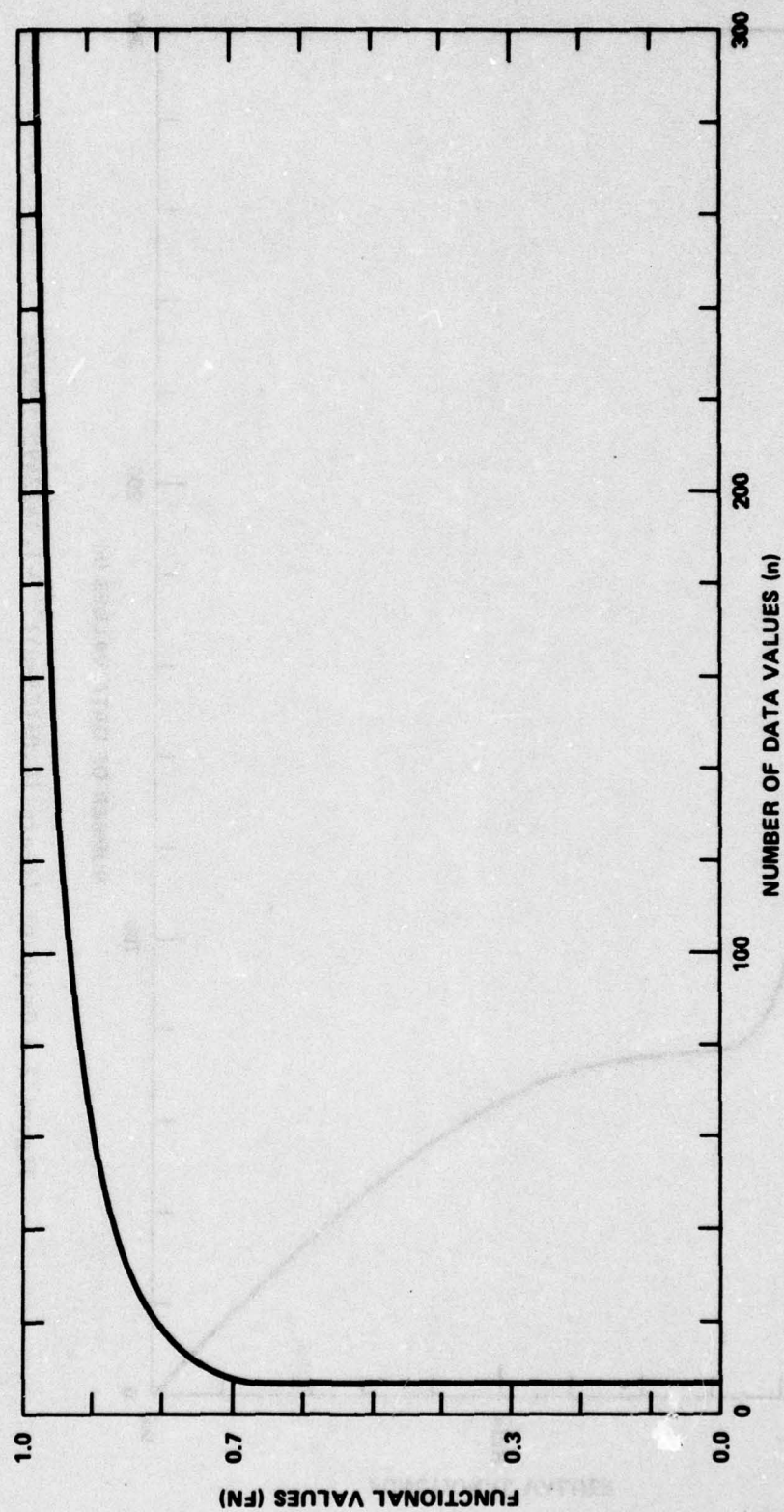


Figure 1 - Graph of Equation (9)

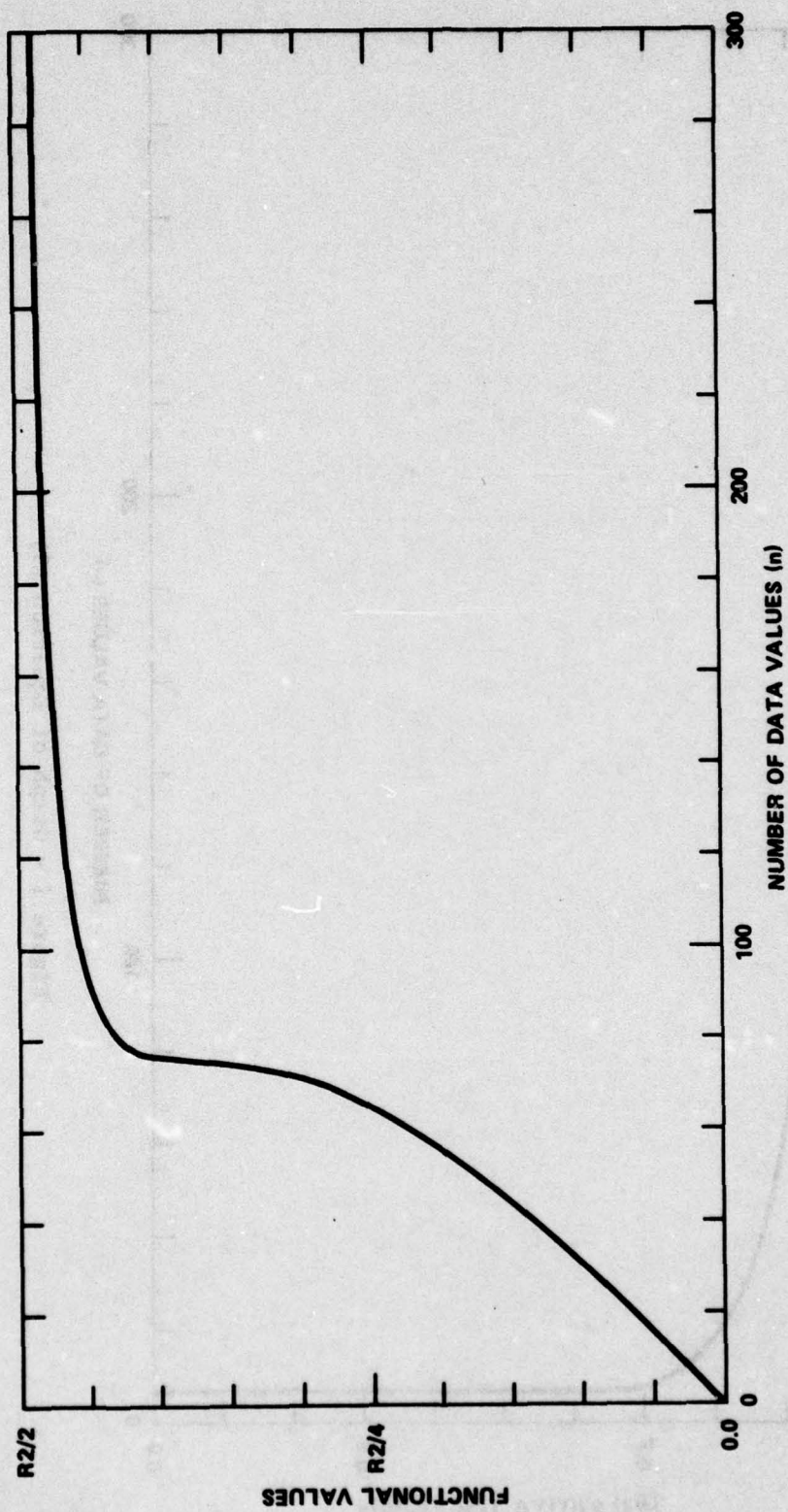


Figure 2 - Graph of $[1 - \exp([0.04(75-n)]^{1/3} - 1.44224957))]R2/2$

2.4.1 A Skewness Test

The skewness test in CHECK uses the approximate distribution of a transformation of the sample skewness. The transformation was introduced by Johnson.⁷ For

$$b_1 = n^{1/2} \sum_{i=1}^n (X_i - \bar{X})^3 / \left[\sum_{i=1}^n (X_i - \bar{X})^2 \right]^{3/2} \quad (12)$$

and

$$W = \left(-1 + \sqrt{\frac{6(n^2 + 27n - 70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}} - 2 \right)^{1/2} \quad (13)$$

D'Agostino⁸ proved that the distribution of

$$Z = R \ln[(y/a)^{1/2} + (1 + y/a)^{1/2}] \quad (14)$$

is approximately standard normal (i.e., with mean zero and variance one) if X_1, X_2, \dots, X_n is a random sample from a normal distribution and

$$R = (\ln W)^{-1/2} \quad (15)$$

$$y = [b_1^2 (n+1)(n+3)] / [6(n-2)] \quad (16)$$

$$a = 2/(W^2 - 1) \quad (17)$$

This test can be used only on samples with at least eight data values.

To determine the significance level corresponding to a particular value of Z, the following approximation (see Johnson and Kotz,⁴ page 55) to the standard normal distribution function is used:

$$\text{Prob}[X \leq x] \approx 1 - 0.5(1 + 0.196854x + 0.115194x^2 + 0.000344x^3 + 0.0019527x^4)^{-4} \quad (18)$$

for $x \geq 0$ and X having a standard normal distribution.

2.4.2 The Shapiro-Wilk Test

If X_1, X_2, \dots, X_n is a random sample arranged in ascending order, the Shapiro-Wilk test statistic is given by

$$W = \left(\sum_{i=1}^n a_{in} X_i \right)^2 / \sum_{i=1}^n (X_i - \bar{X})^2 \quad (19)$$

where the vector $a_n = (a_{1n}, \dots, a_{nn})$ is given by

$$a_n = m'V^{-1}/(m'V^{-1}V^{-1}m)^{1/2} \quad (20)$$

and where m is the vector of expected values, and V the covariance matrix, of the order statistics of a sample of size n from a standard normal distribution. The properties of this statistic were first examined by Shapiro and Wilk,⁹ who tabulated the values a_{in} for n from 2 through 50. Among the properties of W , if the X_j 's are from a normal distribution with any mean and variance, is that the value W has a lower bound of

$$E_n = na_{1n}^2/(n-1) \quad (21)$$

and an upper bound of one. In CHECK the a_{in} 's are in an array called A and the E_n 's are in an array called $EPSILN$. The significance level for a given value of W is found by applying the approximation, given in

Equation (18), of the standard normal distribution to the following transformation of W:

$$ZA = D_n \ln[(W-E_n)/(1-W)] - G_n \quad (22)$$

The values of D_n and G_n are in arrays DELTA and GAMMA, respectively, in CHECK.

2.4.3 The Transformed D Statistic Test

If X_1, X_2, \dots, X_n is a random sample arranged in ascending order, the D statistic is given by

$$D = \frac{\sum_{i=1}^n [1 - (n+1)/2] X_i}{\left(n^3 \sum_{i=1}^n (X_i - \bar{X})^2 \right)^{1/2}} \quad (23)$$

For values of n from 50 to 1000 for various significance levels, D'Agostino⁶ has tabulated the critical points of the following transformation of D

$$Y = \left(D - \frac{1}{2\sqrt{\pi}} \right) \sqrt{\frac{24\pi n}{12\sqrt{3} - 37 + 2}}$$

$$= \frac{\sqrt{n} (D - 0.28209479)}{0.02998598} \quad (24)$$

under the assumption that the random sample is from an arbitrary normal distribution. A curve-fitting routine was used to derive functions for the 5 percent, 10 percent, 90 percent, and 95 percent significance levels.

CHECK uses these functions to determine, for a given sample size, the critical points corresponding to these four significance levels. A judgment concerning the normality of the data is then based on a comparison of the sample value for Y, the transformed D statistic, with these four critical points.

2.5 THE NORMAL BOUNDS CALCULATOR

2.5.1 Calculation of Bounds

If X_1, X_2, \dots, X_n is a random sample from a normal distribution, then for

$$g = [(n+1)/n]^{1/2} t(1-p/2) \quad (25)$$

where $t(1-p/2)$ is the $(1-p/2)$ critical point of a Students' t distribution with $n-1$ degrees of freedom, the probability that another independent draw from the same normal distribution will lie between $X-gs$ and $X+gs$ is $1-p$ (see Johnson and Kotz,⁴ p. 74). The same result holds if

$$g^2 = [(n+1)/n] F(1-p) \quad (26)$$

where $F(1-p)$ is the $(1-p)$ critical point of an F distribution with one and $n-1$ degrees of freedom. CHECK uses the IMSL³ routine MDFD to determine the significance level for $ng^2/(n+1)$.

2.5.2 Use Decision

This paragraph describes the criteria used in CHECK to decide whether or not a set of bounds based on a normal assumption is to be recommended, and if so, which set. Any recommended bounds would be included on TAPE2 (see subsection 3.3). The criteria presented here are based on the

professional judgment of the author and the requirements of the application for which CHECK was first written.

The decision to recommend normal bounds or not is based on the results of the skewness test discussed in paragraph 2.4.1 and, depending on the size of the data set, on the results of either the Shapiro-Wilk test (see paragraph 2.4.2) or the transformed D statistic test (see paragraph 2.3.4). As in the exponential bounds calculator (see paragraph 2.3.3), CHECK uses the value of PCIND to determine whether normal bounds are to be recommended.

If the number of data values is less than eight, the skewness test cannot be used and normal bounds are recommended only if the significance level of the Shapiro-Wilk test is at least 0.5. For larger samples, normal bounds are recommended if the significance of the skewness test is at least 0.9 and either the significance level of the Shapiro-Wilk test is at least 0.48 (50 data values or less) or the D statistic lies between the 10 percent and 90 percent significance values (more than 50 data values). Alternatively, normal bounds will be recommended if the significance of the skewness test is at least 0.75, and either the significance level of the Shapiro-Wilk test is at least 0.5 or the D statistic lies between the 5 percent and 95 percent significance values. As in the exponential case, any recommended bounds will be such that the expected probability is at least 0.9775 that another independent draw from the same presumed normal distribution will lie within the bounds.

2.6 THE RANGE BOUNDS CALCULATOR

If X_1, X_2, \dots, X_n is a random sample from any continuous distribution arranged in ascending order, the expected value of the probability P that some future independent draw from the same distribution will lie between X_s and X_t , where $s < t$, is given by

$$E(P) = (t-s)/(n+1) \quad (27)$$

(see Wilks¹⁰). In particular, the expected probability that some future draw will be between X_1 and X_n is $(n-1)/(n+1)$.

P is actually a random variable. As n increases, the expected value of P increases rapidly. However, the probability that P is within 0.005 of its expected value (i.e., accuracy in the second decimal) starts low and increases slowly. Figure 3 exemplifies these statements. Some of the detail lost in Figure 3 is summarized in Table 1. Note from Table 1 that the expected value of P is already 0.75 with just seven data points, but the probability that the actual value of P would round to 0.75 is only 2.5 percent. Similarly, E(P) is 0.9 at nineteen data points, but the probability that the actual value would round to 0.90 is only 5.7 percent. The probability that P is within 0.005 of its expected value doesn't reach 90 percent until the number of data points is about 390. By that time the expected value of P is nearly 0.995.

Because of the requirements of the application for which CHECK was first written, X_1 and X_n are included on TAPE2 (see subsection 3.3) as a suggested set of range bounds for each data set analyzed.

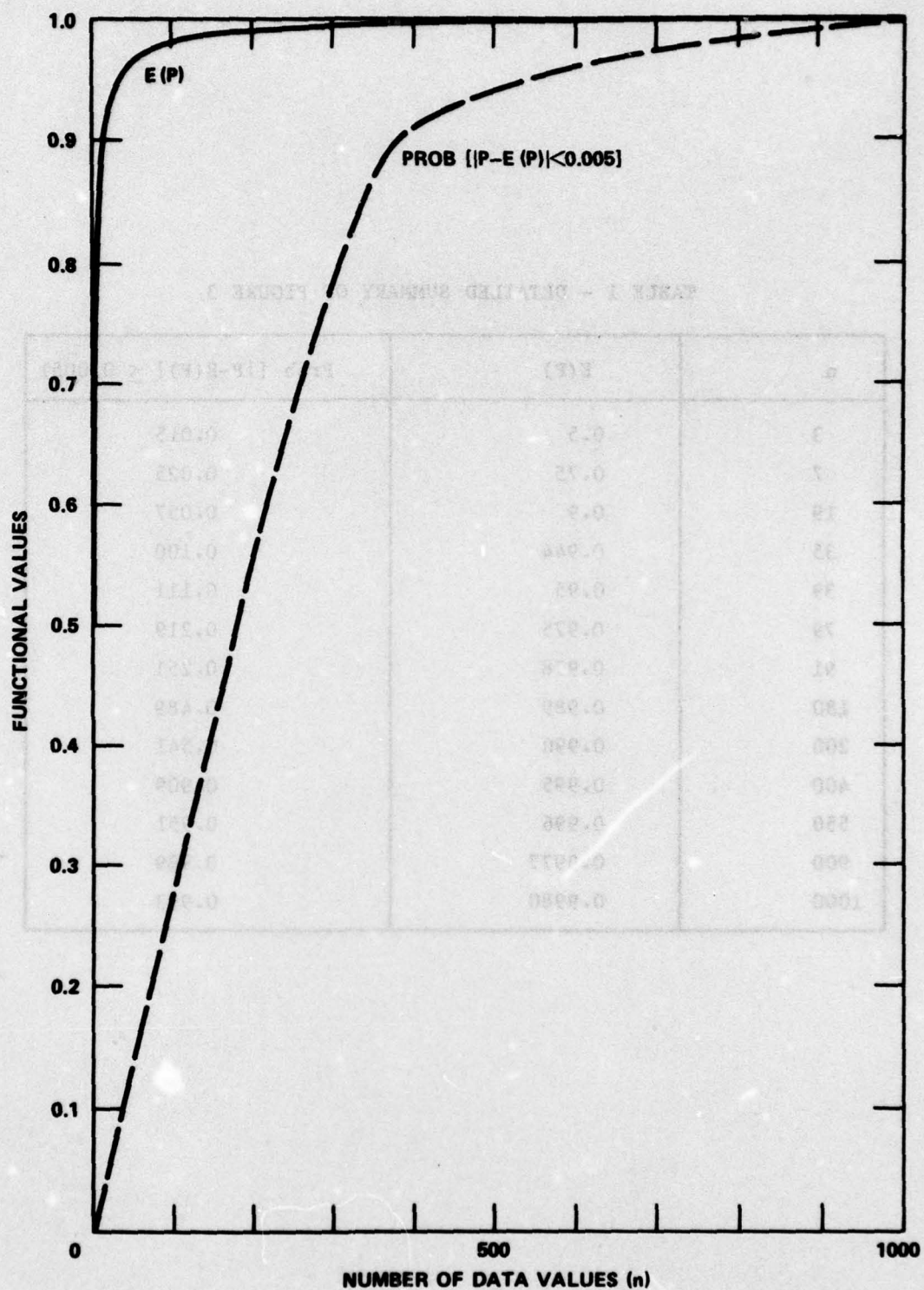


Figure 3 - Properties of $P = \text{Prob } [X_1 \leq X \leq X_n]$

TABLE 1 - DETAILED SUMMARY OF FIGURE 3

n	E(P)	Prob $[P-E(P) \leq 0.005]$
3	0.5	0.015
7	0.75	0.025
19	0.9	0.057
35	0.944	0.100
39	0.95	0.111
79	0.975	0.219
91	0.978	0.251
180	0.989	0.489
200	0.990	0.541
400	0.995	0.909
550	0.996	0.951
900	0.9977	0.989
1000	0.9980	0.993

3. USE OF CHECK

3.1 INPUT

The input to CHECK is a series of data sets to be examined. Each data set has the following form:

- o Title
- o Data
- o End indication

The title must be contained in the first thirty columns of the first record of the data set. Each data value must be contained on a separate record and within the first thirty columns of its record. The data value is considered complete at the first blank after at least one digit (see subsection 2.2). The end-of-data-set indicator is a string of ten or more minus signs and/or decimal points contained in the first thirty columns of a record. No data value or title may be contained on the same record as the end-of-data-set indicator.

As presented in Appendix B, CHECK reads all input from a file called TAPE3. This may be an attached permanent file or a local file created prior to execution of the program. If it is desired to read the input from cards, the PROGRAM card (the first card in CHECK) should be changed to replace "TAPE3," with "TAPE3=INPUT," (see the listing of CHECK in Appendix B). Sample input is given in Appendix C.

3.2 MECHANICS OF USE

Use of CHECK on the CDC 6000 computers at DTNSRDC is implemented through the following set of control cards:

<u>CARD</u>	<u>Description</u>
Job card	Standard
Charge card	Standard
FTN.	Loads CHECK
ATTACH,TAPE3.	Attaches the data set
ATTACH,IMSL.	Attaches the IMSL library

LDSET,LIB=IMSL.

Loads the IMSL Library.
This is necessary since the
IMSL routine MDFD is used.

LGO.

Executes the program

If CHECK is provided on a permanent file (say, one called CHECK) rather than in card form, the FTN card can be replaced by

ATTACH,TEMP,CHECK,...

FTN(I=TEMP)

If a compiled version of CHECK is available on a permanent file (again one called CHECK), the FTN card may be replaced by

ATTACH,LGO,CHECK,...

If the input data are provided on cards rather than on a permanent file, and if changing CHECK as specified in subsection 3.1 is undesirable, the ATTACH,TAPE3,... card may be replaced by

COPYBF(INPUT,TAPE3)

REWIND,TAPE3.

If the summary information on TAPE2 (see subsection 3.3) is to be cataloged as a permanent file,

REQUEST,TAPE2,*PF.

should be included after the FTN card and

CATALOG,TAPE2,...

should be included after the LGO card. If this summary information is to be printed instead, the following cards should be included after the LGO card:

REWIND,TAPE2.

COPYSBF(TAPE2,OUTPUT)

3.3 OUTPUT

The output of CHECK is in two parts, printed and file. The printed output for each data set consists of

- o The data set title
- o Error messages
- o The number of blank fields
- o The data values in ascending order
- o The number of data values
- o The average of the data values
- o The standard deviation of the data values
- o The results of the large sample test for exponentiality (see paragraph 2.3.1)
- o The results of the uniformity test for exponentiality (see paragraph 2.3.2)
- o An indication of the bounds to be used if an exponential assumption is to be made, including a table of such bounds (printed only if an exponential assumption is not unrealistic)
- o The results of the skewness test for normality (see paragraph 2.5.1) if the number of data values is at least eight
- o The results of the Shapiro-Wilk test for normality (see paragraph 2.5.2) if the number of data values is no larger than fifty
- o The result of the transformed D statistic test for normality (see paragraph 2.5.3) if the number of data values is larger than fifty
- o An indication of the bounds to be used if a normal assumption is to be made, including a table of such bounds (printed only if a normal assumption is not unrealistic)
- o One or more sets of range bounds

The file output is a summary of the printed output. Included in the file output for each data set are a line for exponential bounds if the three tests discussed in subsection 2.3 indicate that an exponential assumption is reasonable, a line for normal bounds if the tests discussed in subsection 2.5 indicate that a normal assumption is reasonable, and a line for range bounds. Each line of the file output contains

- o The data set title
- o The number of data points
- o The average of the data values
- o The standard deviation of the data values
- o Suggested lower and upper bounds
- o Expected probability that a future value would lie within the bounds, given this distributional assumption
- o An indication of the distributional assumption (E for exponential, N for normal, R for range or no distributional assumption)

File output from CHECK is provided in the form of a local file called TAPE2. Disposition of TAPE2 is discussed in subsection 3.2.

Samples of both printed and file output are given in Appendix C.

APPENDIX A

GLOSSARY

CDC	Control Data Corporation
DEC	Digital Equipment Corporation
DTNSRDC	David W. Taylor Naval Ship Research and Development Center
exp	Exponential function; $\exp(a) = e^a$
IMSL	International Mathematical and Statistical Library
ln	Natural (base e) logarithm function
MDFD	IMSL statistical subroutine
MRC	Maintenance Requirement Card
NAVSEC	Naval Ship Engineering Center
pdf	Probability density function
SMMS	Shipsystem Maintenance Monitoring and Support
SST	SMMS Site Team
standard normal distribution	Normal distribution with mean zero and variance one

APPENDIX B
PROGRAM LISTING

```

PROGRAM CHECK(INPUT,OUTPUT,TAPE6=OUTPUT,TAPE3,TAPE2)
C   USE THIS ONE FOR GENERAL DATA SETS
C   REPORT DATA WILL BE GENERATED
C   DIMENSION X(1200),C(14),FLD(30),TITL(3)
C   DATA (C(I),I=1,14)/1H0,1H1,1H2,1H3,1H4,1H5,1H6,1H7,1H8,1H9,
11H.,1H-,1H,1H+/
C   REWIND 3
C   N = 0
C$$$CHANGE THE FOLLOWING CARD AS NECESSARY TO MATCH THE DIMENSION
C$$$OF X.
C   NX=1200
C   WRITE(6,3000)
C   NUM=0
C *** TITLE READER ***
10 READ(3,5003) (TITL(I),I=1,3)
C   IF(EOF(3)) 420,20
20 NUM = NUM + 1
C   WRITE(6,3003) NUM
C   WRITE(6,5001) (TITL(I),I=1,3)
C   N=0
C   NBFLD=0
C *** DATA READER ***
30 READ(3,5000) (FLD(I),I=1,30)
C   NCOL=0
C   N=N+1
C   NBLANK=0
C   NEND=0
C   SIGN=1.0
C   BUILD=0.0
C   WHOLE=10.0
C   FR=1.0
C   FRACT=1.0
C   IDG=0
C *** DATA INTERPRETER ***
40 NCOL=NCOL+1
C   CCOL=FLD(NCOL)
C   IF ALL COLUMNS HAVE BEEN READ, GO TO 70
C   IF(NCOL.GT.30) GO TO 70
C   IF CURRENT COLUMN IS BLANK, GO TO 120
C   IF(CCOL.EQ.C(13)) GO TO 120
C   IF CURRENT COLUMN HAS MINUS, GO TO 110
C   IF(CCOL.EQ.C(12)) GO TO 110
C   IF CURRENT COLUMN HAS DECIMAL, GO TO 90
C   IF(CCOL.EQ.C(11)) GO TO 90
C   CHECK TO SEE IF CURRENT COLUMN HAS A DIGIT
C   DO 50 I=1,10
C   IF(CCOL.NE.C(I)) GO TO 50

```

```

NEND=0
IDB=1
FRACT=FRACT*FR
BUILD = BUILD*WHOLE + FRACT*FLOAT(I-1)
GO TO 40
50 CONTINUE
C      IF CURRENT COLUMN HAS PLUS, GO TO 40
      IF(CCOL.EQ.C(14)) GO TO 40
60 WRITE(6,5005) (FLD(I),I=1,30)
      WRITE(6,5002)
      N=N-1
      GO TO 30
C      BUILD THE NUMBER
70 BUILD = BUILD * SIGN
      IF(NBLANK.GE.30) GO TO 80
      IF(IDG.EQ.0) GO TO 60
      IF(N.LE.NX) X(N)=BUILD
      GO TO 30
80 NBFLD=NBFLD+1
      N=N-1
      GO TO 30
C      TRANSFER FROM INTEGER PART TO FRACTIONAL PART
90 IF(WHOLE.LT.5.0) GO TO 100
      WHOLE=1.0
      FR=0.1
100 NEND=NEND+1
      IF(NEND.GE.10) GO TO 130
      GO TO 40
C      MAKE THE NUMBER NEGATIVE
110 SIGN=-1.0
      NEND=NEND+1
      IF(NEND.GE.10) GO TO 130
      GO TO 40
C      COUNT THE BLANKS
120 NBLANK=NBLANK+1
      IF(IDG)40,40,70
C *** EXPONENTIAL CHECKER ***
130 N=N-1
      IF(N.LT.3) GO TO 380
      IF(N.GT.NX) GO TO 370
      M=N/2
      IF(M.LT.1) M=1
      X1=0.0
      X2=0.0
      IF(N.GT.2*M) X2=X(N)
      DO 140 I=1,M
      X1=X1+X(I)
140 X2=X2+X(I+M)
      X1=(X1/FLOAT(M))
      K=N-M
      X2=(X2/FLOAT(K))
      Q=FLOAT(N)

```

```

CALL WSTEST(X,N,Y,P,Z,AVE,STDEV,R2)
IF(NBFLD.NE.0) WRITE(6,4013) NBFLD
WRITE(6,4000) (X(I),I=1,N)
WRITE(6,4001) N,AVE,STDEV
IF(STDEV.LE.0.0) GO TO 390
C      LARGE SAMPLE TEST FOR EXPONENTIALITY
WRITE(6,4022) R2
PCIND=0.0
IF(N.LE.5) GO TO 150
FN=AMIN1(0.30255*ALOG(Q)-1.8237537*Q**0.09+2.2946,0.98)
PCIND=-1.0
SGN=-1.0
IF(N.LT.75) SGN=1.0
IF(R2.GE.FN) PCIND=(1.0-EXP(SGN*(0.04*ABS(FLOAT(N-75))))**((1.0/3.0)
1-1.44224957))*0.5*R2+0.05*(R2-FN)/(1.0-FN)
150 X1=X1-X(1)
X2=X2-X(1)
F=0.0
IF(ABS(X2).GT.0.0000001) F=X1/X2
M=2*M
K=2*K
CALL NDFD(F,M,K,PF,IER)
IF(IER.GE.128) PF=0.0
IF(PF.GT.0.5) PF=1.0-PF
PF=2.0*PF
PCIND=PCIND+PF
C      UNIFORMITY TEST FOR EXPONENTIALITY
IF(PF.LT.0.1) GO TO 220
IF(PF-0.5) 160,170,170
160 WRITE(6,4007)
GO TO 180
170 WRITE(6,4008)
180 WRITE(6,4009) X(1),AVE,PF
C *** EXPONENTIAL BOUNDS CALCULATOR ***
190 DIFF=AVE-X(1)
WRITE(6,4027) X(1),DIFF,X(1)
COMP=0.0
BL=X(1)
IF(AMAX1(DIFF,STDEV).GE.1.8*AMIN1(DIFF,STDEV)) PCIND=0.0
DO 210 K=10,200
T=0.1*FLOAT(K)
PP=1.0-(1.0/(1.0+(T/Q)))**Q
IF(PP.LE.COMP) GO TO 210
BU=T*DIFF+BL
COMP=PP+0.005
WRITE(6,4003) T,PP,BL,BU
IF(COMP.LE.0.9825.OR.PCIND.LT.0.5) GO TO 200
PP=100.0*PP
C      PROVIDE EXPONENTIAL BOUNDS ON TAPE2
WRITE(2,3501) (TITL(I),I=1,3),N,AVE,STDEV,BL,BU,PP
PCIND=0.0
200 IF(COMP.GE.1.0) GO TO 230
210 CONTINUE

```

```

        GO TO 230
220  WRITE(6,4010) PF,IER
      IF(IER.NE.0) GO TO 230
      IF(R2.GT.0.85.AND.N.GT.50) GO TO 190
      IF(PCIND.GE.0.5.AND.N.LE.50) GO TO 190
C *** NORMAL CHECKER ***
230  PCIND=0.3
      PZ=1.0
      IF(N.LT.8) GO TO 240
C      SKEWNESS TEST FOR NORMALITY
      Z=ABS(Z)
      PCIND=0.0
      PZ=1.0/(1.0+Z*(.196854+Z*(.115194+Z*(.000344+Z*.019527))))**4
      WRITE(6,4015) PZ
      IF(PZ.GE.0.75) PCIND=PCIND+0.3
      IF(PZ.GT.0.9) PCIND=PCIND+0.15
      IF((PZ-0.5)*(PZ-0.1).LT.0.0) WRITE(6,4004)
      IF(PZ.GE.0.5) WRITE(6,4005)
      IF(PZ.LE.0.1) WRITE(6,4006)
      IF(N.GT.50) GO TO 250
C      SHAPIRO-WILK TEST FOR NORMALITY
240  WRITE(6,4014) P
      IF(P.GE.0.5) PCIND=PCIND+0.3
      IF(P.GE.0.48) PCIND=PCIND+0.15
      IF((P-0.5)*(P-0.1).LT.0.0) WRITE(6,4004)
      IF(P.GE.0.5) WRITE(6,4005)
      IF(P.LE.0.1) WRITE(6,4006)
      GO TO 310
C      TRANSFORMED D STATISTIC TEST FOR NORMALITY
250  P10=-1.3439065+.199408805/EXP(.01*Q)-39.1133537/Q+.247836076/EXP(Q
      1*.02)+634.84298/Q**2
      P=1.0
      IF(Y.LT.P10) GO TO 280
      P90=1.21938688+.205750184/EXP(Q*.01)-47.655998/Q+.240379393/EXP(Q
      1*.02)+701.334854/Q**2
      IF(Y.GT.P90) GO TO 260
      WRITE(6,4016) Y,P10,P90
      PCIND=PCIND+0.3
      GO TO 300
260  P95=1.55358157+.30118663/EXP(Q*.01)-67.593821/Q+.362900004/EXP(Q*
      102)+1016.22535/Q**2
      IF(Y.GE.P95) GO TO 270
      WRITE(6,4017) Y,P90,P95
      PCIND=PCIND+0.15
      GO TO 300
270  WRITE(6,4018) Y,P95
      P=0.0
      GO TO 300
280  P05=-1.73657547+.29836828/EXP(Q*.01)-59.054124/Q+.34545809/EXP(Q*
      1.02)+977.76416/Q**2
      IF(Y.LE.P05) GO TO 290
      WRITE(6,4019) Y,P10,P05

```

```

PCIND=PCIND+0.15
GO TO 300
290 WRITE(6,4020) Y,P05
P=0.0
300 IF((Y-0.5)*(Y-0.1).LT.0.0) WRITE(6,4004)
IF(Y.GE.0.5) WRITE(6,4005)
IF(Y.LE.0.1) WRITE(6,4006)
C *** NORMAL BOUNDS CALCULATOR ***
310 IF(P.LT.0.1.OR.PZ.LT.0.25) GO TO 340
WRITE(6,4002)
ADDON=FLOAT(N)/FLOAT(N+1)
M=N-1
G=0.4
COMP=0.0
DO 330 K=5,200
G=G+0.1
FP=G*G*ADDON
CALL MDFFD(FP,1,M,PP,IER)
IF(IER.GE.128.OR.PP.LE.COMP) GO TO 330
BL=Ave-G*STDEV
BU=Ave+G*STDEV
COMP=PP+0.005
WRITE(6,4003) G,PP,BL,BU
IF(COMP.LE.0.985.OR.PCIND.LT.0.5) GO TO 320
PP=100.0*PP
C PROVIDE NORMAL BOUNDS ON TAPE2
WRITE(2,3502) (TITL(I),I=1,3),N,Ave,STDEV,BL,BU,PP
PCIND=0.0
320 IF(COMP.GE.1.0) GO TO 340
330 CONTINUE
C *** RANGE BOUNDS CALCULATOR ***
340 WRITE(6,4011)
L=1
J=N
COMP=2.0
BL=X(1)-1.0
BU=X(N)+1.0
DO 360 I=1,N
IF(L.GE.J) GO TO 10
PP=FLOAT(J-L)/(Q+1.0)
IF(COMP-PP.LT.0.005.OR.X(L).LE.BL.OR.X(J).GE.BU) GO TO 350
BL=X(L)
BU=X(J)
COMP=PP
WRITE(6,4021) BL,BU,PP
PP=100.0*PP
C PROVIDE RANGE BOUNDS ON TAPE2
IF(I.EQ.1) WRITE(2,3503) (TITL(L),L=1,3),N,Ave,STDEV,BL,BU,PP
350 L=L+1
360 J=J-1
GO TO 10

```

```

C *** ERROR STATEMENTS ***
370 WRITE(6,3002) N,NX
    GO TO 10
380 WRITE(6,4012) N
    IF(NBFLD.NE.0) WRITE(6,4013) NBFLD
    IF(N-1) 10,400,410
390 PP=100.0*FLOAT(N-1)/FLOAT(N+1)
    WRITE(2,3503) (TITL(I),I=1,3),N,AVE,STDEV,AVE,AVE,PP
    GO TO 10
400 WRITE(2,3504) (TITL(I),I=1,3),X(1)
    WRITE(6,4000) X(1)
    GO TO 10
410 AVE=AMAX1(X(1),X(2))
    X(1)=AMIN1(X(1),X(2))
    X(2)=AVE
    AVE=0.5*(X(1)+X(2))
    WRITE(2,3505) (TITL(I),I=1,3),AVE,X(1),X(2)
    WRITE(6,4000) (X(I),I=1,N)
    GO TO 10
420 STOP
3000 FORMAT(1H1,* THIS PROGRAM IS TO DETERMINE IF A GIVEN SET OF*
1,* DATA FITS NORMAL OR*/ EXPONENTIAL ASSUMPTIONS.*)
3002 FORMAT(/** THIS PROBLEM CANNOT BE ATTACKED AT THIS TIME.*/
1* THE TOTAL NUMBER OF DATA VALUES,*,I6,*, EXCEEDS THE DIMENSION OF
2 X, WHICH IS *,I6/* IF YOU RERUN WITH X REDIMENSIONED, BE SURE TO
3 ALSO CHANGE THE VALUE OF NX.*/)
3003 FORMAT(////////* -----PROBLEM *,I3,*-----*/)
3501 FORMAT(3A10,1X,I4,1X,2(F10.3,1X),2(F8.3,1X),F4.1,* E*)
3502 FORMAT(3A10,1X,I4,1X,2(F10.3,1X),2(F8.3,1X),F4.1,* N*)
3503 FORMAT(3A10,1X,I4,1X,2(F10.3,1X),2(F8.3,1X),F4.1,* R*)
3504 FORMAT(3A10,4X,*1*,1X,F10.3)
3505 FORMAT(3A10,4X,*2*,1X,F10.3,12X,2(F8.3,1X),5X,*R*)
4000 FORMAT(* DATA:*,6(1X,F10.3))
4001 FORMAT(/5X,*NUMBER OF POINTS: *,I5/5X,*AVERAGE VALUE (M): *,F16.6
1/5X,*STANDARD DEVIATION (S): *,F16.6)
4002 FORMAT(5X,*USE (M-KS) AS A LOWER BOUND AND (M+KS) AS AN UPPER BOUN
1D*/5X,*WHERE THE EXPECTED PROPORTION OF POPULATION FOUND WITHIN TH
2E LIMITS*/5X,*IS A FUNCTION OF K AS SPECIFIED IN THE FOLLOWING TAB
3LE*/6X,*K*,6X,*PROPORTION OF POPULATION*,5X,*LOWER BOUND*,5X,
4*UPPER BOUND*/)
4003 FORMAT(4X,F4.1,12X,F9.6,12X,E13.7,3X,E13.7)
4004 FORMAT(5X,*NORMAL ASSUMPTION QUESTIONABLE.*)
4005 FORMAT(5X,*NORMAL ASSUMPTION HAS SOME MERIT.*)
4006 FORMAT(5X,*NORMAL ASSUMPTION DOUBTFUL.*)
4007 FORMAT(/5X,*EXPONENTIAL ASSUMPTION QUESTIONABLE WITH*)
4008 FORMAT(/5X,*EXPONENTIAL ASSUMPTION PROMISING WITH*)
4009 FORMAT(5X,*A=*,F16.6,5X,*AND MEAN=*,F16.6/5X,*SIGNIFICENCE LEVEL
1IS *,F9.6)
4010 FORMAT(/3X,*EXPONENTIAL ASSUMPTION DOUBTFUL SINCE SIGNIFICENCE LEV
1EL IS *,F9.6/5X,*(MDFD ERROR CODE *,I3,*))
4011 FORMAT(///3X,*IF NORMAL AND EXPONENTIAL ASSUMPTIONS CANNOT BE MAD
1E OR ARE TOO RISKY,*)

```

```

4012 FORMAT(/5X,*THE NUMBER OF DATA POINTS,*,18,*, IS TOO SMALL.*
      1/5X,*THE PROBLEM IS NOT TACKLED.*)
4013 FORMAT(// * THE NUMBER OF BLANK FIELDS IS *,I6/)
4014 FORMAT(/5X,*THE SHAPIRO-WILK SIGNIFICANCE LEVEL IS *,F12.6)
4015 FORMAT(/5X,*SIGNIFICANCE LEVEL OF SKEWNESS TEST IS *,F12.6)
4016 FORMAT(/5X,*THE VALUE,*,E13.6,*, FOR THE TRANSFORMED D STATISTIC
      1LIES WITHIN THE*/5X,*80 PERCENT LIMITS OF *,E13.6,* AND *,E13.6)
4017 FORMAT(/5X,*THE VALUE,*,E13.6,*, FOR THE TRANSFORMED D STATISTIC
      1LIES ABOVE THE*/5X,*80 PERCENT LIMIT,*,E13.6,*, BELOW THE 90 PERCE
      2NT LIMIT,*,E13.6)
4018 FORMAT(/5X,*THE VALUE,*,E13.6,*, FOR THE TRANSFORMED D STATISTIC
      1LIES ABOVE THE*/5X,*90 PERCENT LIMIT OF *,E13.6)
4019 FORMAT(/5X,*THE VALUE,*,E13.6,*, FOR THE TRANSFORMED D STATISTIC
      1LIES BELOW THE*/5X,*80 PERCENT LIMIT,*,E13.6,*, ABOVE THE 90 PERCE
      2NT LIMIT,*,E13.6)
4020 FORMAT(/5X,*THE VALUE,*,E13.6,*, FOR THE TRANSFORMED D STATISTIC
      1LIES BELOW THE*/5X,*90 PERCENT LIMIT OF *,E13.6)
4021 FORMAT(/5X,*USE *,E16.6,* AS A LOWER BOUND AND *,E16.6,* AS AN*
      1/5X,*UPPER BOUND. EXPECTED PROPORTION OF POPULATION BOUNDED IS*
      2,F9.6)
4022 FORMAT(/5X,*SIGNIFICANCE LEVEL OF LARGE SAMPLE TEST FOR EXPONENTI
      1ALITY IS*,F9.6/5X,*THIS RESULT SHOULD NOT BE USED IF THE NUMBER O
      2F DATA POINTS IS LESS THAN*/5X,*30. ITS SIGNIFICANCE SHOULD BE TA
      3KEN WITH A GRAIN OF SALT IF THE NUMBER*/5X,*OF DATA POINTS IS LESS
      4 THAN 100. FOR 100 POINTS OR MORE*/5X,*EXPONENTIALITY IS DOUBTFUL
      5 IF THE SIGNIFICANCE LEVEL IS LESS*/5X,*THAN .98 AND PROMISING IF
      6IT EXCEEDS .99.*/)
4027 FORMAT(/5X,*USE*,F16.6,5X,*AS THE LOWER BOUND AND */
      25X,F16.6,*K + *,F16.6,5X,*AS THE UPPER BOUND*/5X,*THE EXPECTED PRO
      3PORTION OF THE POPULATION FALLING WITHIN THESE*/5X,*BOUNDS DEPENDS
      4 UPON K AS INDICATED IN THE FOLLOWING TABLE*/6X,*K*,6X,*PROPORTIO
      5N OF POPULATION*,5X,*LOWER BOUND*,5X,*UPPER BOUND*/)
5000 FORMAT(30A1)
5001 FORMAT(5X,3A10//)
5002 FORMAT(* THE NUMBER IN THIS FIELD IS NOT INCLUDED IN THE DATA SET.
      1*/)
5003 FORMAT(3A10)
5005 FORMAT(* CONTENTS OF FIELD: *,30A1)
      END

```

```

      SUBROUTINE WSTEST(X,N,Y,P,Z,AVE,STDEV,R2)
C --- SUBROUTINE TO APPLY THE W TEST FOR NORMALITY (COMPLETE SAMPLES)
C      OF SHAPIRO AND WILK
C      X IS A SET OF DATA TO BE TESTED FOR NORMALITY
C      N IS THE NUMBER OF ITEMS IN ARRAY A
C      X WILL BE SORTED IN INCREASING ORDER X(1).LE.X(2).LE.X(3)...
C      RESULT W IS THE TEST STATISTIC OF SHAPIRO AND WILK
C      P IS THE SIGNIFICANCE LEVEL ASSOCIATED WITH W
C      RESTRICTIONS 3.LE.N.LE.50
C      DIMENSION X(N)

```

DIMENSION INDEXA(51),A(625),GAMMA(50),DELTA(50),EPSILN(50)
 C INDEXA(N) IS THE INDEX OF THE FIRST ITEM IN ARRAY *A* TO BE USED
 DATA(INDEXA(I),I=2,51) / 1, 2, 3, 5, 7, 10, 13, 17, 21
 1, 26, 31, 37, 43, 50, 57, 65, 73, 82, 91, 101, 111, 122, 133, 145, 157
 2, 170, 183, 197, 211, 226, 241, 257, 273, 290, 307, 325, 343, 362, 381, 401, 421
 3, 442, 463, 485, 507, 530, 553, 577, 601, 626/
 DATA(A(I),I=1,210)/.7071,.7071,.6872,.1677,.6646,.2413,.6431,.2806
 1,.0875,.6233,.3031,.1401,.6052,.3164,.1743,.0561,.5888,.3244,.1976
 2,.0947,.5739,.3291,.2141,.1224,.0399,.5601,.3315,.2260,.1429,.0695
 3,.5475,.3325,.2347,.1586,.0922,.0303,.5359,.3325,.2412,.1707,.1099
 4,.0539,.5251,.3318,.2460,.1802,.1240,.0727,.0240,.5150,.3306,.2495
 5,.1878,.1353,.0880,.0433,.5056,.3290,.2521,.1939,.1447,.1005,.0593
 6,.0196,.4968,.3273,.2540,.1988,.1524,.1109,.0725,.0359,.4886,.3253
 7,.2553,.2027,.1587,.1197,.0837,.0496,.0163,.4808,.3232,.2561,.2059
 8,.1641,.1271,.0932,.0612,.0303,.4734,.3211,.2565,.2085,.1686,.1334
 9,.1013,.0711,.0422,.0140,.4643,.3185,.2578,.2119,.1736,.1399,.1092
 A,.0804,.0530,.0263,.4590,.3156,.2571,.2131,.1764,.1443,.1150,.0878
 B,.0618,.0368,.0122,.4542,.3126,.2563,.2139,.1787,.1480,.1201,.0941
 C,.0696,.0459,.0228,.4493,.3098,.2554,.2145,.1807,.1512,.1245,.0997
 D,.0764,.0539,.0321,.0107,.4450,.3069,.2543,.2148,.1822,.1539,.1283
 E,.1046,.0823,.0610,.0403,.0200,.4407,.3043,.2533,.2151,.1836,.1563
 F,.1316,.1089,.0876,.0672,.0476,.0284,.0094,.4366,.3018,.2522,.2152
 G,.1848,.1584,.1346,.1128,.0923,.0728,.0540,.0358,.0178,.4328,.2992
 H,.2510,.2151,.1857,.1601,.1372,.1162,.0965,.0778,.0598,.0424,.0253
 I,.0084,.4291,.2968,.2499,.2150,.1864,.1616,.1395,.1192,.1002,.0822
 J,.0650,.0483,.0320,.0159/
 DATA(A(I),I=211,420) /.4254,.2944,.2487,.2148,.1870,.1630,.1415
 1,.1219,.1036,.0862,.0697,.0537,.0381,.0227,.0076,.4220,.2921,.2475
 2,.2145,.1874,.1641,.1433,.1243,.1066,.0899,.0739,.0585,.0435,.0289
 3,.0144,.4188,.2898,.2463,.2141,.1878,.1651,.1449,.1265,.1093,.0931
 4,.0777,.0629,.0485,.0344,.0206,.0068,.4156,.2876,.2451,.2137,.1880
 5,.1660,.1463,.1284,.1118,.0961,.0812,.0669,.0530,.0395,.0262,.0131
 6,.4127,.2854,.2439,.2132,.1882,.1667,.1475,.1301,.1140,.0988,.0844
 7,.0706,.0572,.0441,.0314,.0187,.0062,.4096,.2834,.2427,.2127,.1883
 8,.1673,.1487,.1317,.1160,.1013,.0873,.0739,.0610,.0484,.0361,.0239
 9,.0119,.4068,.2813,.2415,.2121,.1883,.1678,.1496,.1331,.1179,.1036
 A,.0900,.0770,.0645,.0523,.0404,.0287,.0172,.0057,.4040,.2794,.2403
 B,.2116,.1883,.1683,.1505,.1344,.1196,.1056,.0924,.0798,.0677,.0559
 C,.0444,.0331,.0220,.0110,.4015,.2774,.2391,.2110,.1881,.1686,.1513
 D,.1356,.1211,.1075,.0947,.0824,.0706,.0592,.0481,.0372,.0264,.0158
 E,.0053,.3989,.2755,.2380,.2104,.1880,.1689,.1520,.1366,.1225,.1092
 F,.0967,.0848,.0733,.0622,.0515,.0409,.0305,.0203,.0101,.3964,.2737
 G,.2368,.2098,.1878,.1691,.1526,.1376,.1237,.1108,.0986,.0870,.0759
 H,.0651,.0546,.0444,.0343,.0244,.0146,.0049,.3940,.2719,.2357,.2091
 I,.1876,.1693,.1531,.1384,.1249,.1123,.1004,.0891,.0782,.0677,.0575
 J,.0476,.0379,.0283,.0188,.0094/
 DATA(A(I),I=421,625) /.3917,.2701,.2345,.2085,.1874,.1694,.1535
 1,.1392,.1259,.1136,.1020,.0909,.0804,.0701,.0602,.0506,.0411,.0318
 2,.0227,.0136,.0045,.3894,.2684,.2334,.2078,.1871,.1695,.1539,.1398
 3,.1269,.1149,.1035,.0927,.0824,.0724,.0628,.0534,.0442,.0352,.0263
 4,.0175,.0087,.3872,.2667,.2323,.2072,.1868,.1695,.1542,.1405,.1278
 5,.1160,.1049,.0943,.0842,.0745,.0651,.0560,.0471,.0383,.0296,.0211

```

6,.0126,.0042,.3850,.2651,.2313,.2065,.1865,.1695,.1545,.1410,.1286
7,.1170,.1062,.0959,.0860,.0765,.0673,.0584,.0497,.0412,.0328,.0245
8,.0163,.0081,.3830,.2635,.2302,.2058,.1862,.1695,.1548,.1415,.1293
9,.1180,.1073,.0972,.0876,.0783,.0694,.0607,.0522,.0439,.0357,.0277
A,.0197,.0118,.0039,.3808,.2620,.2291,.2052,.1859,.1695,.1550,.1420
B,.1300,.1189,.1085,.0986,.0892,.0801,.0713,.0628,.0546,.0465,.0385
C,.0307,.0229,.0153,.0076,.3789,.2604,.2281,.2045,.1855,.1693,.1551
D,.1423,.1306,.1197,.1095,.0998,.0906,.0817,.0731,.0648,.0568,.0489
E,.0411,.0335,.0259,.0185,.0111,.0037,.3770,.2589,.2271,.2038,.1851
F,.1692,.1553,.1427,.1312,.1205,.1105,.1010,.0919,.0832,.0748,.0667
G,.0588,.0511,.0436,.0361,.0288,.0215,.0143,.0071,.3751,.2574,.2260
H,.2032,.1847,.1691,.1554,.1430,.1317,.1212,.1113,.1020,.0932,.0846
I,.0764,.0685,.0608,.0532,.0459,.0386,.0314,.0244,.0174,.0104,.0035
J/

```

```

DATA (GAMMA(I),I=3,50) /0.625,1.107,1.530,2.010,2.356,2.696,2.968
1,3.262,3.485,3.731,3.936,4.155,4.373,4.567,4.713,4.885,5.018,5.153
2,5.291,5.413,5.508,5.605,5.704,5.803,5.905,5.988,6.074,6.160,6.248
3,6.324,6.402,6.480,6.559,6.640,6.721,6.803,6.887,6.961,7.035,7.111
4,7.188,7.266,7.345,7.414,7.484,7.555,7.615,7.677/

```

```

DATA (DELTA(I),I=3,50) /0.386,0.714,0.935,1.138,1.245,1.333,1.400
1,1.471,1.515,1.571,1.613,1.655,1.695,1.724,1.739,1.770,1.786,1.802
2,1.818,1.835,1.848,1.862,1.876,1.890,1.905,1.919,1.934,1.949,1.965
3,1.976,1.988,2.000,2.012,2.024,2.037,2.049,2.062,2.075,2.088,2.101
4,2.114,2.128,2.141,2.155,2.169,2.183,2.198,2.212/

```

```

DATA (EPSILN(I),I=3,50) /7.500,.6297,.5521,.4963,.4533,.4186,.3900
1,.3660,.3451,.3270,.3111,.2969,.2842,.2727,.2622,.2528,.2440,.2359
2,.2264,.2207,.2157,.2106,.2063,.2020,.1980,.1943,.1907,.1872,.1840
3,.1811,.1781,.1755,.1727,.1702,.1677,.1656,.1633,.1612,.1591,.1572
4,.1552,.1534,.1516,.1499,.1482,.1466,.1451,.1436/

```

C *** CALCULATE THE W NORMALITY TEST STATISTIC OF SHAPIRO AND WILK

C *** NON-NORMAL RETURN FOR N .LT. 3

W =0.

P =0.

IF (N.LT.3) RETURN

C *** SORT SAMPLE IN ASCENDING ORDER

DO 930 I=2,N

JMAX =I-1

YTEST =X(I)

IF (YTEST.GE.X(JMAX)) GO TO 930

DO 900 IJ=1,JMAX

IF (YTEST.LT.X(IJ)) GO TO 910

900 CONTINUE

910 DO 920 IK=IJ,JMAX

II =IJ+JMAX-IK

920 X(II+1) =X(II)

X(IJ) =YTEST

930 CONTINUE

S=0.0

R2=0.0

DO 940 I = 1,N

R2=R2+ALOG(FLOAT(I))

940 S=S+X(I)

```

Q=FLOAT(N)
AVE=S/Q
R2=R2/Q
SSQ=0.0
Y3=0.0
Y4=0.0
Y6=0.0
Y7=0.0
Y5=Q+1.
Y2=-Y5/2.0
DO 950 I=1,N
Y2=Y2+1.0
Y4=Y4+Y2*X(I)
Y8=X(I)-AVE
SSQ=SSQ+Y8*Y8
Y3=Y3+Y8*Y8*Y8
Y5=Y5-1.0
Y9=ALOG(Y5)-R2
Y6=Y6+Y9*Y9
950 Y7=Y7+Y9*Y8
C *** SSQ      =SYMMETRIC ESTIMATE OF VARIANCE
C *** NON-NORMAL RETURN IF SSQ = 0.
      U        =1.
      P        =1.
      IF       (SSQ.LE.0.) RETURN
C *** R2      = R-SQUARED FOR LARGE SAMPLE TEST FOR EXPONENTIALITY
      R2=Y7*Y7/(SSQ*Y6)
      IF(N.GT.50) GO TO 970
C *** B      =WEIGHTED SUM OF ORDER STATISTICS
      KMIN    =INDEXA(N)
      KMAX    =INDEXA(N+1)-1
      II      =1
      IJ      =N
      B       =0.
      DO 960 IK=KMIN,KMAX
      B       =B+A(IK)*(X(IJ)-X(II))
      II      =II+1
960 IJ      =IJ-1
      BSQ     =B*B
C *** W      = TEST STATISTIC
      W       =BSQ/SSQ
C *** NON-NORMAL RETURN IF W .LE. EPISLN
      AHOLD   =(W-EPISLN(N))/(1.-W)
      P       =0.
      IF      (AHOLD.LE.0.) GO TO 980
C *** P      =SIGNIFICANCE LEVEL ASSOCIATED WITH W
      ZA      =DELTA(N)*ALOG(AHOLD)-GAMMA(N)
      ZZ      =ABS(ZA)
      ZZ2     =ZZ**2
      P       =1./(2.*(1.+196854*ZZ+.115194*ZZ2
      +.000344*ZZ*ZZ2+.019527*ZZ2*ZZ2)**4)
1

```

```

      IF      (ZA.GT.0.) P =1.-P
      GO TO 980
C *** CALCULATE STATISTICS FOR D TEST OF NORMALITY
970 D=Y4/SQRT(FLOAT(N*N*N)*SSQ)
      Y=SQRT(FLOAT(N))*(D-.28209479)/.02998598
C *** CALCULATE STATISTICS FOR SKEWNESS TEST
980 IF(N.LT.8) GO TO 990
      B1=SQRT(FLOAT(N))*Y3/SQRT(SSQ*SSQ*SSQ)
      Y1=FLOAT((N*N+27*N-70)*(N+1)*(N+3))/FLOAT((N-2)*(N+5)*(N+7)*(N+9))
      W2=-1.0+SQRT(6.0*Y1-2.0)
      YA=B1*B1*FLOAT((N+1)*(N+3))*(W2-1.0)/(12.0*FLOAT(N-2))
      R=1.0/SQRT(ALOG(SQRT(W2)))
      Z=R*ALOG(SQRT(YA)+SQRT(YA+1.0))
990 STDEV=SQRT(SSQ/FLOAT(N-1))
      RETURN
      END

```

APPENDIX C

SAMPLE RUN

This appendix provides a sample run of CHECK. The data used in the sample run are

```

TEST DATA SET 1
- .9
-.888
6
X - .933
-1.
-.7
.12- VALUE 4
156VALUE 5
-1 VALUE 1
2 VALUE 2
-0.3 VALUE 3
0
-.854
1
0./
0.0
-.49
.....
TEST DATA SET 2
0
0
0
+ 1
+ 2
+ 3
+ 1
+ 2
+ 1
-----

```

Because the "2" in the title of the second data set is in the thirty-first column of the data card, it will not be printed in the title in the output. Note also that, in the thirteenth data field of the first data set, a slash occurs prior to the blank which indicates the number has been

completed. Because the slash is an inadmissible character, the data field is rejected (see the sample printed output beginning at the bottom of this page). The fourth and eighth data fields in the first data set are also rejected because the "X" and the "V" are also inadmissible characters.

The output consists of both printed and file output, as discussed in subsection 3.3. The file output for the sample run is

TEST DATA SET 1	14	.196	1.873	-1.000	4.383 98.0 E
TEST DATA SET 1	14	.196	1.873	-1.000	6.000 86.7 R
TEST DATA SET	9	1.111	1.054	0.000	3.000 80.0 R

The printed output for the sample run completes this appendix. Note that the difference between the two largest values in the first data set is large compared, for example, with the standard deviation. When the analyst examines this printed output, he may wish to delete the largest value as a possible "outlier", or bad value, and run CHECK again with the resulting data set.

-----PROBLEM 1-----

TEST DATA SET 1

CONTENTS OF FIELD: X - .933
THE NUMBER IN THIS FIELD IS NOT INCLUDED IN THE DATA SET.

CONTENTS OF FIELD: 156VALUE 5
THE NUMBER IN THIS FIELD IS NOT INCLUDED IN THE DATA SET.

CONTENTS OF FIELD: 0./
THE NUMBER IN THIS FIELD IS NOT INCLUDED IN THE DATA SET.

DATA:	-1.000	-1.000	-.900	-.888	-.854	-.700
	-.490	-.300	-.120	0.000	0.000	1.000
	2.000	6.000				

NUMBER OF POINTS: 14
AVERAGE VALUE (M): .196286
STANDARD DEVIATION (S): 1.873355

SIGNIFICANCE LEVEL OF LARGE SAMPLE TEST FOR EXPONENTIALITY IS .847145
 THIS RESULT SHOULD NOT BE USED IF THE NUMBER OF DATA POINTS IS LESS THAN
 30. ITS SIGNIFICANCE SHOULD BE TAKEN WITH A GRAIN OF SALT IF THE NUMBER
 OF DATA POINTS IS LESS THAN 100. FOR 100 POINTS OR MORE
 EXPONENTIALITY IS DOUBTFUL IF THE SIGNIFICANCE LEVEL IS LESS
 THAN .98 AND PROMISING IF IT EXCEEDS .99.

EXPONENTIAL ASSUMPTION PROMISING WITH
 A= -1.000000 AND MEAN= .196286
 SIGNIFICANCE LEVEL IS .993696

USE -1.000000 AS THE LOWER BOUND AND
 1.196286K + -1.000000 AS THE UPPER BOUND
 THE EXPECTED PROPORTION OF THE POPULATION FALLING WITHIN THESE
 BOUNDS DEPENDS UPON K AS INDICATED IN THE FOLLOWING TABLE

K	PROPORTION OF POPULATION	LOWER BOUND	UPPER BOUND
1.0	.619360	-.1000000E+01	.1962857E+00
1.1	.653171	-.1000000E+01	.3159143E+00
1.2	.683785	-.1000000E+01	.4355429E+00
1.3	.711522	-.1000000E+01	.5551714E+00
1.4	.736669	-.1000000E+01	.6748000E+00
1.5	.759481	-.1000000E+01	.7944286E+00
1.6	.780190	-.1000000E+01	.9140571E+00
1.7	.798999	-.1000000E+01	.1033686E+01
1.8	.816095	-.1000000E+01	.1153314E+01
1.9	.831642	-.1000000E+01	.1272943E+01
2.0	.845790	-.1000000E+01	.1392571E+01
2.1	.858671	-.1000000E+01	.1512200E+01
2.2	.870407	-.1000000E+01	.1631829E+01
2.3	.881104	-.1000000E+01	.1751457E+01
2.4	.890861	-.1000000E+01	.1871086E+01
2.5	.899765	-.1000000E+01	.1990714E+01
2.6	.907896	-.1000000E+01	.2110343E+01
2.7	.915324	-.1000000E+01	.2229971E+01
2.8	.922113	-.1000000E+01	.2349600E+01
2.9	.928323	-.1000000E+01	.2469229E+01
3.0	.934006	-.1000000E+01	.2588857E+01
3.1	.939208	-.1000000E+01	.2708486E+01
3.3	.948341	-.1000000E+01	.2947743E+01
3.5	.956020	-.1000000E+01	.3187000E+01
3.7	.962488	-.1000000E+01	.3426257E+01
3.9	.967948	-.1000000E+01	.3665514E+01
4.2	.974602	-.1000000E+01	.4024400E+01
4.5	.979798	-.1000000E+01	.4383286E+01
4.9	.985026	-.1000000E+01	.4861800E+01
5.5	.990333	-.1000000E+01	.5579571E+01
6.6	.995516	-.1000000E+01	.6895486E+01

SIGNIFICANCE LEVEL OF SKEWNESS TEST IS .000419
NORMAL ASSUMPTION DOUBTFUL.

THE SHAPIRO-WILK SIGNIFICANCE LEVEL IS .000062
NORMAL ASSUMPTION DOUBTFUL.

IF NORMAL AND EXPONENTIAL ASSUMPTIONS CANNOT BE MADE OR ARE TOO RISKY,

USE -.100000E+01 AS A LOWER BOUND AND .600000E+01 AS AN
UPPER BOUND. EXPECTED PROPORTION OF POPULATION BOUNDED IS .866667

USE -.900000E+00 AS A LOWER BOUND AND .100000E+01 AS AN
UPPER BOUND. EXPECTED PROPORTION OF POPULATION BOUNDED IS .600000

USE -.888000E+00 AS A LOWER BOUND AND 0. AS AN
UPPER BOUND. EXPECTED PROPORTION OF POPULATION BOUNDED IS .466667

USE -.700000E+00 AS A LOWER BOUND AND -.120000E+00 AS AN
UPPER BOUND. EXPECTED PROPORTION OF POPULATION BOUNDED IS .200000

USE -.490000E+00 AS A LOWER BOUND AND -.300000E+00 AS AN
UPPER BOUND. EXPECTED PROPORTION OF POPULATION BOUNDED IS .066667

-----PROBLEM 2-----

TEST DATA SET

THE NUMBER OF BLANK FIELDS IS 3

DATA:	0.000	0.000	0.000	1.000	1.000	1.000
	2.000	2.000	3.000			

NUMBER OF POINTS: 9
AVERAGE VALUE (M): 1.11111
STANDARD DEVIATION (S): 1.054093

EXPONENTIAL ASSUMPTION DOUBTFUL SINCE SIGNIFICENCE LEVEL IS .010049
(NDFD ERROR CODE 0)

SIGNIFICENCE LEVEL OF SKEWNESS TEST IS .431168
NORMAL ASSUMPTION QUESTIONABLE.

THE SHAPIRO-WILK SIGNIFICENCE LEVEL IS .192172
NORMAL ASSUMPTION QUESTIONABLE.

USE (M-KS) AS A LOWER BOUND AND (M+KS) AS AN UPPER BOUND
WHERE THE EXPECTED PROPORTION OF POPULATION FOUND WITHIN THE LIMITS
IS A FUNCTION OF K AS SPECIFIED IN THE FOLLOWING TABLE

K	PROPORTION OF POPULATION	LOWER BOUND	UPPER BOUND
.5	.352066	.5840648E+00	.1638157E+01
.6	.415164	.4786556E+00	.1743567E+01
.7	.474687	.3732463E+00	.1848976E+01
.8	.530347	.2678371E+00	.1954385E+01
.9	.581962	.1624278E+00	.2059794E+01
1.0	.629445	.5701856E-01	.2165204E+01
1.1	.672802	-.4839070E-01	.2270613E+01
1.2	.712114	-.1538000E+00	.2376022E+01
1.3	.747526	-.2592092E+00	.2481431E+01
1.4	.779234	-.3646185E+00	.2586841E+01
1.5	.807467	-.4700277E+00	.2692250E+01
1.6	.832480	-.5754370E+00	.2797659E+01
1.7	.854539	-.6808462E+00	.2903068E+01
1.8	.873912	-.7862555E+00	.3008478E+01
1.9	.890865	-.8916647E+00	.3113887E+01
2.0	.905650	-.9970740E+00	.3219296E+01
2.1	.918510	-.1102483E+01	.3324705E+01
2.2	.929666	-.1207893E+01	.3430115E+01
2.3	.939324	-.1313302E+01	.3535524E+01
2.4	.947671	-.1418711E+01	.3640933E+01
2.5	.954874	-.1524120E+01	.3746342E+01
2.6	.961082	-.1629530E+01	.3851752E+01
2.7	.966429	-.1734939E+01	.3957161E+01
2.9	.974987	-.1945757E+01	.4167980E+01
3.1	.981317	-.2156576E+01	.4378798E+01
3.4	.987862	-.2472804E+01	.4695026E+01
3.8	.993068	-.2894441E+01	.5116663E+01
4.8	.998135	-.3948533E+01	.6170755E+01

IF NORMAL AND EXPONENTIAL ASSUMPTIONS CANNOT BE MADE OR ARE TOO RISKY,

USE 0. AS A LOWER BOUND AND .300000E+01 AS AN
UPPER BOUND. EXPECTED PROPORTION OF POPULATION BOUNDED IS .800000

USE .100000E+01 AS A LOWER BOUND AND .100000E+01 AS AN
UPPER BOUND. EXPECTED PROPORTION OF POPULATION BOUNDED IS .200000

PROPORTION OF POPULATION	LOWER BOUND	UPPER BOUND
0.0	0.000000E+00	0.000000E+00
0.1	0.000000E+00	0.000000E+00
0.2	0.000000E+00	0.000000E+00
0.3	0.000000E+00	0.000000E+00
0.4	0.000000E+00	0.000000E+00
0.5	0.000000E+00	0.000000E+00
0.6	0.000000E+00	0.000000E+00
0.7	0.000000E+00	0.000000E+00
0.8	0.000000E+00	0.000000E+00
0.9	0.000000E+00	0.000000E+00
1.0	0.000000E+00	0.000000E+00
1.1	0.000000E+00	0.000000E+00
1.2	0.000000E+00	0.000000E+00
1.3	0.000000E+00	0.000000E+00
1.4	0.000000E+00	0.000000E+00
1.5	0.000000E+00	0.000000E+00
1.6	0.000000E+00	0.000000E+00
1.7	0.000000E+00	0.000000E+00
1.8	0.000000E+00	0.000000E+00
1.9	0.000000E+00	0.000000E+00
2.0	0.000000E+00	0.000000E+00
2.1	0.000000E+00	0.000000E+00
2.2	0.000000E+00	0.000000E+00
2.3	0.000000E+00	0.000000E+00
2.4	0.000000E+00	0.000000E+00
2.5	0.000000E+00	0.000000E+00
2.6	0.000000E+00	0.000000E+00
2.7	0.000000E+00	0.000000E+00
2.8	0.000000E+00	0.000000E+00
2.9	0.000000E+00	0.000000E+00
3.0	0.000000E+00	0.000000E+00
3.1	0.000000E+00	0.000000E+00
3.2	0.000000E+00	0.000000E+00
3.3	0.000000E+00	0.000000E+00
3.4	0.000000E+00	0.000000E+00
3.5	0.000000E+00	0.000000E+00
3.6	0.000000E+00	0.000000E+00
3.7	0.000000E+00	0.000000E+00
3.8	0.000000E+00	0.000000E+00
3.9	0.000000E+00	0.000000E+00
4.0	0.000000E+00	0.000000E+00

APPENDIX D

CROSS-REFERENCE TABLE OF VARIABLES AND PROGRAM USAGE

This appendix provides a listing of the variable names and indicates the use of each variable in CHECK.

A	Table of coefficients for the Shapiro-Wilk test for normality.
ADDON	Used in calculation of significance level for normal bounds (see Equation (26)); $ADDON = N/(N+1)$.
AHOLD	Intermediate value used in calculation of the significance level of the Shapiro-Wilk test statistic. $AHOLD = (W - E_n)/(1-W)$ of Equation (22)
AVE	Average of the data values.
B	Weighted sum of the order statistics used in calculating the Shapiro-Wilk test statistic (see BSQ).
BL	Lower bound for the printed lines in the three bounds calculators (see subsections 2.4, 2.6, and 2.7).
BSQ	Square of the weighted sum of the order statistics used in calculating the Shapiro-Wilk test statistic. Numerator of Equation (19).
BU	Upper bound for the printed lines in the three bounds calculators (see subsections 2.4, 2.6, and 2.7)
BUILD	Used to hold the current number as it is being built (see subsection 2.2).
B1	Sample skewness (see Equation (12)).
C	Table of admissible characters (see subsection 2.2).
CCOL	Contents of current column; $CCOL = FLD(NCOL)$.
COMP	Used to control printed output in the bounds calculators. A line will be printed only if PP (see PP) has increased by at least 0.005 since the last printed line.
D	D test statistic (see Equation (21)).
DELTA	Table of values, one of which is used in each Shapiro-Wilk test to calculate the significance level associated with the test statistic (see W). $DELTA(n)$ is the D_n of Equation (22).
DIFF	Difference between the average of the data values (see AVE) and the smallest of the data values (see X).

EPSILN	Table of values, one of which is used in each Shapiro-Wilk test as an initial indication of non-normality of the data. EPSILN(n) is E_n of Equation (21).
F	Uniformity test statistic; f of Equation (3).
FLD	Array containing the current data field.
FN	Minimum function of N to aid in determining the value of PCIND. If R2 is less than FN, an exponential line will not be included on TAPE2 for that data set. See Equation (10)
FP	Used to control printed output; F(1-p) of Equation (26).
FR	Fractional part indicator. FR = 1 before the first decimal point is encountered. FR = 0.1 afterward.
FRACT	Fractional part multiplier. FRACT = (0.1)**i when the i^{th} decimal place is added to BUILD
G	Used to control printed output. G = g of Equations (25) and (26).
GAMMA	Table of values one of which is used in each Shapiro-Wilk test to calculate the significance level associated with the test statistic (see W). GAMMA(n) = G_n of Equation (22).
I	Index.
IDG	Digit indicator. IDG = 0 until a digit is encountered. Then IDG = 1.
IER	IMSL error indicator whose value is assigned by MDFD. If IER is larger than 128, the value of PF has no meaning.
II	Index.
IJ	Index.
IK	Index.
INDEXA	Array of indices for use with A (see A). INDEX (N) is the index of the first item in A to be used in calculation of the Shapiro-Wilk test statistic (see W) when there are N data values.
J	Index of upper bound in range bounds.
JMAX	Limiting value for IJ in sorting the order statistics.
K	N - M (see N and M).
KMAX	Index of the last value in A (see A) to be used in calculation of W (see W).

KMIM	Index of the first value in A (see A) to be used in calculation of W (see W).
L	Index of lower bound in range bounds.
M	Largest integer no larger than one-half of N. If N is even, $M = N/2$. If N is odd, $M = (N-1)/2$.
N	Number of data points.
NBFLD	Number of all-blank data fields in the current data set.
NBLANK	Number of blank columns in the current data field.
NCOL	Number of the column under consideration.
NEND	Number of decimal points and minus signs since the last digit in the current field. NEND = 10 indicates the end of the data set.
NUM	Problem (data set) number.
NX	Dimension of X. If the dimension of X is changed, the value of NX must be changed appropriately.
P	Significance level of the Shapiro-Wilk test statistic.
PCIND	Variable used to control selection of summary information for inclusion on TAPE2 (see paragraphs 2.3.3 and 2.5.2).
PF	Significance level for the uniformity test (see paragraph 2.3.2).
PP	Expected probability that another draw from the same distribution will lie within the suggested bounds (see subsections 2.4, 2.6, and 2.7).
PZ	Significance level for the skewness test statistics (see paragraph 2.5.1)
P05	Five percentile for the transformed D statistic (see paragraph 2.5.3).
P10	Ten percentile for the transformed D statistic (see paragraph 2.5.3).
P90	Ninety percentile for the transformed D statistic (see paragraph 2.5.3).
P95	Ninety-five percentile for the transformed D statistic (see paragraph 2.5.3).
Q	Number of data points; $Q = N$.
R	R of Equations (14) and (15).
R2	Significance level (R^2) of the large sample test for exponentiality (see Equation (1)).
S	Sum of the data values.

SGN	Indicator variable used in calculation of PCIND. SGN = 1 if N is less than 75. Otherwise, SGN = - 1.
SIGN	Sign of the current data value (plus or minus one).
SSQ	Sum of the squared deviations from their mean of the data values.
STDEV	Standard deviation of the data values.
T	Used to control printed output. T = t of Equation (8).
TITL	Data set title.
W	Shapiro-Wilk test statistic (see Equation (19)).
WHOLE	Integer part multiplier. WHOLE = 10 before the first decimal point is encountered for each field. WHOLE = 1 afterward.
W2	Square of W of Equation (13).
X	Array used to store the values in the current data set. In subroutine WSTEST, the values in X are arranged in ascending order so that X(1) is the smallest of the values. (See NX.)
X1	Numerator of the test statistic of the uniformity test for exponentiality (see Equation (3)).
X2	Denominator of the test statistic of the uniformity test for exponentiality (see Equation (3)).
Y	Transformed D statistic (see Equation (24)).
YA	(y/a) of Equation (14)
YTEST	Dummy variable used in sorting the values into ascending order.
Y1	Dummy variable used in calculation of W2 (see W2).
Y2	$i - (N+1)/2$ used in calculation of D (see Equation (23)).
Y3	Sum of $(X(i) - AVE)^3$ used in calculation of B1 (see Equation (11)).
Y4	Sum of $[i - (N+1)/2]X(i)$ used in calculation of D (see Equation (23)).
Y5	$(N+1-j)$ used in calculation of R2 (see Equation (1)).
Y6	Sum of $Y9^2$ used in calculation of R2 (see Y9).
Y7	Sum of $Y8 \times Y9$ used in calculation of R2 (see Y8 and Y9).
Y8	$X(i) - AVE$ used in calculation of SSQ, B1, and R2 (see SSQ and Equations (1) and (12)).
Y9	Difference between $\ln(N+1-i)$ and the average of the $\ln(N+1-j)$'s, used in calculation of R2 (see Equation (12)).
Z	Skewness test statistic (see Equation (14)).

ZA	Intermediate value in the calculation of P (see Equation (22)).
ZZ	Absolute value of ZA (see ZA), used in the calculation of P.
ZZ2	Square of ZA (see ZA), used in the calculation of P.

REFERENCES

1. Epstein, B., "Tests for the Validity of the Assumption that the Underlying Distribution of Life is Exponential. Part I," *Technometrics*, vol. 2, no. 1, pp. 88-101 (1960).
2. Johnston, J., "Econometric Methods," Second Edition, McGraw-Hill Book Co., New York (1972), Chapter 2.
3. "IMSL Library 3 Reference Manual," Fifth Edition, International Mathematical and Statistical Libraries, Houston, Texas (Nov. 1975).
4. Johnson, N. and S. Kotz, "Distributions in Statistics: Continuous Univariate Distributions - 1," Houghton Mifflin Co., Boston (1970).
5. Shapiro, S.S. "A Comparative Study of Various Tests for Normality," *American Statistical Association Journal*, vol. 63, pp. 1343-1372 (1968).
6. D'Agostino, R.B., "An Omnibus Test of Normality for Moderate and Large Size Samples," *Biometrika*, vol. 58, no. 2, pp. 341-348 (1971).
7. Johnson, N.L., "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika*, vol. 36, pp. 149-176 (1949).
8. D'Agostino, R.B., "Transformation to Normality of the Null Distribution of g_1 ," *Biometrika*, vol. 57, no. 3, pp. 679-681 (1970).
9. Shapiro, S.S. and M.B. Wilk, "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, vol. 52, pp. 591-611 (1965).
10. Wilks, S.S., "Determination of Sample Sizes for Setting Tolerance Limits," *Annals of Mathematical Statistics*, vol. 12, pp. 91-96 (1941).

INITIAL DISTRIBUTION

Copies

3	NAVPGSCOL
	1 Library
	1 D. Gaver, Code 55Gv
	1 D. Barr, Code 55Bn
1	NAVSEA, Code 03F (B. Orleans)
5	NAVSEC 6107E1
1	CEL, Code L62 (W. Pierpoint)
12	DDC
1	MSC, Code M-62a5 (P. Morfogenis)

CENTER DISTRIBUTION

Copies	Code	Name
1	18	G. Gleissner
1	1805	E. Cuthill
2	1809.3	D. Harris
1	184.1	H. Feingold
1	187	J. Spurway
1	187	M. Zubkoff
20	187	G. Humfeld
10	5214.1	Reports Distribution
1	522.1	Unclassified Lib (C)
1	522.2	Unclassified Lib (A)

